

The Phase Transition of Matrix Recovery from Gaussian Measurements Matches the Minimax MSE of Matrix Denoising

David L. Donoho^{*} Matan Gavish^{*}
Andrea Montanari^{*†}

February 12, 2013

Abstract

Let X_0 be an unknown M by N matrix. In matrix recovery, one takes $n < MN$ linear measurements y_1, \dots, y_n of X_0 , where $y_i = \text{Tr}(a_i^T X_0)$ and each a_i is a M by N matrix. For measurement matrices with Gaussian i.i.d entries, it is known that if X_0 is of low rank, it is recoverable from just a few measurements. A popular approach for matrix recovery is Nuclear Norm Minimization (NNM): solving the convex optimization problem $\min \|X\|_*$ subject to $y_i = \text{Tr}(a_i^T X)$ for all $1 \leq i \leq n$, where $\|\cdot\|_*$ denotes the nuclear norm, namely, the sum of singular values. Empirical work reveals a *phase transition* curve, stated in terms of the undersampling fraction $\delta(n, M, N) = n/(MN)$, rank fraction $\rho = r/N$ and aspect ratio $\beta = M/N$. Specifically, a curve $\delta^* = \delta^*(\rho; \beta)$ exists such that, if $\delta > \delta^*(\rho; \beta)$, NNM typically succeeds, while if $\delta < \delta^*(\rho; \beta)$, it typically fails.

An apparently quite different problem is matrix denoising in Gaussian noise, where an unknown M by N matrix X_0 is to be estimated based on direct noisy measurements $Y = X_0 + Z$, where the matrix Z has iid Gaussian entries. It has been empirically observed that, if X_0 has low rank, it may be recovered quite accurately from the noisy measurement Y . A popular matrix denoising scheme solves the unconstrained optimization problem $\min \|Y - X\|_F^2/2 + \lambda \|X\|_*$. When optimally tuned, this scheme achieves the asymptotic minimax MSE $\mathcal{M}(\rho) = \lim_{N \rightarrow \infty} \inf_{\lambda} \sup_{\text{rank}(X) \leq \rho N} \text{MSE}(X, \hat{X}_\lambda)$.

^{*}Department of Statistics, Stanford University

[†]Department of Electrical Engineering, Stanford University

We report extensive experiments showing that the phase transition $\delta^*(\rho)$ in the first problem (Matrix Recovery from Gaussian Measurements) coincides with the minimax risk curve $\mathcal{M}(\rho)$ in the second problem (Matrix Denoising in Gaussian Noise): $\delta^*(\rho) = \mathcal{M}(\rho)$, for *any* rank fraction $0 < \rho < 1$.

Our experiments considered matrices belonging to two constraint classes: real M by N matrices, of various ranks and aspect ratios, and real symmetric positive semidefinite N by N matrices, of various ranks. Different predictions $\mathcal{M}(\rho)$ of the phase transition location were used in the two different cases, and were validated by the experimental data.

Contents

1	Introduction	3
2	Methods	6
2.1	Generation of Problem Instances	7
2.2	Convex Optimization	8
2.3	Probability of Exact Recovery	8
2.4	Estimating the Probability of Exact Recovery	8
2.5	Asymptotic Phase Transition Hypothesis	9
2.6	Empirical Phase Transitions	9
2.7	Experimental Design	9
2.8	Computing	10
3	Results	10
4	Discussion: Existing Literature and Our Contributions	13
5	Conclusions	15
A	Asymptotic Minimax MSE Formula	20
B	Summary of Empirical Results	23
C	Data Deposition	28
D	Code Deposition	29

1 Introduction

Let X_0 be an unknown M by N matrix. How many measurements must we obtain in order to ‘completely know’ X_0 ? While it seems that MN measurements must be necessary, in recent years intense research in applied mathematics, optimization and information theory, has shown that, when X_0 is of low rank, we may efficiently recover it from a relatively small number of linear measurements by convex optimization [1–3]. Applications have been developed in fields ranging widely, for example from video and image processing [4], to quantum state tomography [5], to collaborative filtering [1, 6].

Specifically, let $\mathcal{A} : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^n$ be a linear operator and consider measurements $y = \mathcal{A}(X_0)$. If $n < MN$, the problem of inferring X_0 from y may be viewed as attempting to solve an underdetermined system of equations. Under certain circumstances, it has been observed that this (seemingly hopeless) task can be accomplished by solving the so-called nuclear norm minimization problem

$$(P_{nuc}) \quad \min \|X\|_* \quad \text{subject to } y = \mathcal{A}(X). \quad (1)$$

Here the *nuclear norm* $\|X\|_*$ is the sum of singular values of X . For example, it was found that if X_0 is sufficiently low rank, with a principal subspace in a certain sense incoherent to the measurement operator \mathcal{A} , then the solution $X_1 = X_1(y)$ to (P_{nuc}) is precisely X_0 . Such incoherence can be obtained by letting \mathcal{A} be random, for instance if $\mathcal{A}(X_0)_i = \text{Tr}(a_i^T X_0)$ with $a_i \in \mathbb{R}^{m \times n}$ having i.i.d. Gaussian entries. In this case we speak of “*matrix recovery from Gaussian measurements*” [3].

A key phrase from the previous paragraph: ‘if X_0 is *sufficiently* low rank’. Clearly there must be a quantitative trade-off between the rank of X_0 and the number of measurements required, such that higher rank matrices require more measurements. In the Gaussian measurements model, with N sufficiently large, empirical work by Recht, Xu and Hassibi [7, 8], Fazel, Parillo and Recht [3], Tanner and Wei [9] and Oymak and Hassibi [10], documents a *phase transition* phenomenon. For matrices of a given rank, there is a fairly precise number of required samples, in the sense that a transition from non recovery to complete recovery takes place sharply as the number of samples varies across this value. For example, in Figure 1 below we report results obtained in our own experiments, showing that, for reconstructing matrices of size 60 by 60 which are of rank 20, 2600 Gaussian measurements are sufficient with very high probability, but 2400 Gaussian measurements are insufficient with very high probability.

In this paper, we present a simple and explicit formula for the phase transition curve in matrix recovery from Gaussian measurements. The formula

δ	0.67	0.68	0.68	0.68	0.68	0.69	0.69	0.69	0.69	0.70	0.71	0.72	0.73
n	2400	2437	2446	2455	2465	2474	2483	2492	2502	2511	2538	2575	2612
$\hat{\pi}(r n, N)$	0.00	0.00	0.30	0.20	0.30	0.45	0.60	0.80	0.60	0.90	1.00	1.00	1.00

Figure 1: Data from typical Phase Transition experiment. Here $N = 60$, $r = 20$, and the number n of Gaussian measurements varies. Note: our formula predicts an asymptotic phase transition at $\delta^* = 0.6937$, corresponding to $n = 2497$. And, indeed, the success probability is close to $1/2$ at that n . All runs involved $T = 20$ Monte Carlo trials.

arises in an apparently unrelated problem: matrix de-noising in Gaussian noise. In this problem, we again let X_0 denote an M by N matrix, and we observe $Y = X_0 + Z$, where Z is Gaussian iid noise $Z_{ij} \sim \mathcal{N}(0, 1)$. We consider the following nuclear norm de-noising scheme:

$$(P_{nuc, \lambda}) \quad \min \left\{ \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_* \right\}. \quad (2)$$

In this problem the measurements Y are direct, so in some sense complete, but noisy. The solution $\hat{X}_\lambda(Y)$ can be viewed as a shrinkage estimator. In the basis of the singular vectors U_Y and V_Y of Y , the solution $\hat{X}_\lambda(Y)$ is diagonal, and the diagonal entries are produced by soft thresholding of the singular values of Y .

Because the measurements y in the matrix recovery problem are noiseless but incomplete, while the measurements Y in the matrix denoising problem are complete but noisy, the problems seem quite different. Nevertheless, we show here that there is a deep connection between the two problems.

Let us quantify performance in the denoising problem by the minimax MSE, namely

$$\mathcal{M}(\rho; M, N) = \min_{\lambda} \max_{\text{rank}(X) \leq \rho N} \text{MSE}(X_0, \hat{X}_\lambda(Y)),$$

where MSE refers to the dimension-normalized mean-squared error

$$\frac{1}{MN} E \|X - \hat{X}_\lambda\|_F^2$$

and subscript F denotes Frobenius norm. The asymptotic minimax MSE $\mathcal{M}(\rho; \beta) = \lim_{N \rightarrow \infty} \mathcal{M}(\rho; \beta N, N)$ has been derived in [11]. Explicit formulas for the curve $\rho \mapsto \mathcal{M}(\rho; \beta)$ appear in the Appendix. A parametric form is given for the case of asymptotically square matrices, $\beta = 1$. Figures 2 and 3 depict the various minimax MSE curves.

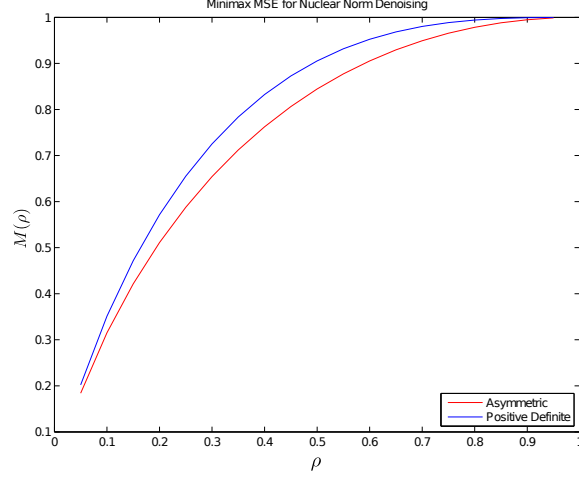


Figure 2: The two asymptotic minimax MSE's $M(\rho|\mathbf{X})$: $\mathbf{X} = M_N$ (black), $\mathbf{X} = \text{Symm}_N$ (red), in the case of square matrices. For non square matrices, see curves in Figure 3

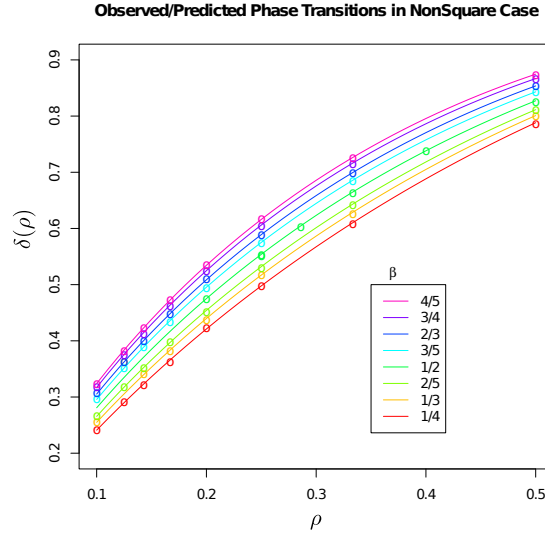


Figure 3: Curves: Asymptotic Minimax MSE's for nonsquare matrix settings $\text{Mat}_{M,N}$; varying shape factor $\beta = M/N$. Points: Empirical phase transition locations for matrix recovery from incomplete measurements, see Table 8.

We can now state our main hypothesis for matrix recovery from Gaussian measurements.

Main Hypothesis: Asymptotic Phase Transition Formula. *Consider a sequence of matrix recovery problems with parameters $\{(r, n, M, N)\}_{N \geq 1}$ having limiting fractional rank $\rho = \lim_{N \rightarrow \infty} r/N$, limiting aspect ratio $\beta = \lim_{N \rightarrow \infty} M/N$, and limiting incompleteness fraction $\delta = \lim_{N \rightarrow \infty} n/(MN)$. In the limit of large problem size N , the solution $X_1(y)$ to the nuclear norm minimization problem (P_{nuc}) is correct with probability converging to one if $\delta > \mathcal{M}(\rho; \beta)$ and incorrect with probability converging to one if $\delta < \mathcal{M}(\rho; \beta)$. In short: The asymptotic phase transition $\delta^*(\rho, \beta)$ in Gaussian matrix recovery is equal to the asymptotic minimax MSE $\mathcal{M}(\rho; \beta)$.*

In particular, for the case of small rank r , by studying the small ρ asymptotics of Eq. (14), we obtain that reconstruction is possible from $n \geq 2r(M + N + \sqrt{MN})(1 + O(r/N))$ measurements, but not from substantially less.

This brief announcement tests this hypothesis by conducting a substantial computer experiment generating large numbers of random problem instances. We use statistical methods to check for disagreement between the hypothesis and the predicted phase transition. To bolster the solidity of our results, we conduct the experiment in two different settings: (1) the matrix X_0 is a general M by N matrix, for various rank fractions ρ and aspect ratios β ; (2) X_0 is a symmetric positive definite matrix, for various rank fractions ρ . In the latter case the positive semidefinite constraint is added to the convex program (P_{nuc}) . As described below, there are different asymptotic MSE curves for the two settings. We demonstrate an empirically accurate match in each of the cases, showing the depth and significance of the connection we expose here.

In the discussion and conclusion we connect our result with related work in the field of sparsity-promoting reconstructions, where the same formal identity between a minimax MSE and a phase transition boundary has been observed, and in some cases even proved. We also discuss recent rigorous evidence towards establishing the above matrix recovery phase transition formula.

2 Methods

We investigated the hypothesis that the asymptotic phase transition boundary agrees with the proposed phase transition formula to within experimental error.

For notational simplicity we will focus here on the case $M = N$, and defer

the case of non-square matrices to the SI. Hence, we will drop throughout the main text the argument $\beta = 1$. The asymptotic phase plane at point (ρ, δ) is associated to triples (r, n, N) , where $\rho = r/N \in [0, 1]$ is the rank fraction, and $\delta = \delta(n, N|\mathbf{X}) = n/\dim(\mathbf{X})$ is the under sampling ratio, where $\dim(\mathbf{X})$ is the dimension of the underlying collection of matrices \mathbf{X} . We performed a sequence of experiments, one for each tuple, in which we generated random rank- r N by N matrices $X_0 \in \mathbf{X}$, random measurement matrices $\mathcal{A} = A$ of size $n \times N^2$, and obtained random problem instances (y, A) . We then applied a convex optimization procedure, obtaining a putative reconstruction $\hat{X} = \hat{X}(y, A)$. We declared a reconstruction successful when the Frobenius norm was smaller than a threshold. Our raw empirical observations consist of a count of empirical successes and sample sizes at a selected set of positions (ρ, δ) and a selected set of problem sizes N . From these raw counts we produce fitted success probabilities $\hat{\pi}(r|n, N, \mathbf{X})$. The finite- N phase transition is the place where the true underlying probability of successful reconstruction take the value 50%. We tested the hypothesis that the finite- N transition was consistent with the proposed asymptotic phase transition formula.

This section discusses details of data generation and data analysis.

2.1 Generation of Problem Instances

Each problem instance (y, A) was generated by, first, generating a random rank r matrix X_0 , then, generating a random measurement matrix $A = A_{n, N^2}$ and then applying $y = A \cdot \text{vec}(X_0)$.

We considered problem instances of two specific types, corresponding to matrices $X_0 \in \mathbf{X}$, with \mathbf{X} one of two classes of matrices

- Mat_N : all $N \times N$ matrices with real-valued entries
- Sym_N : all $N \times N$ real symmetric matrices which are nonnegative-semidefinite

In the case $\mathbf{X} = Mat_N$, we consider low-rank matrices $X_0 = UV'$ where U and V are each N by r partial orthogonal matrices in the Stiefel manifold $St(N, r)$. The matrices are uniformly distributed on $St(N, r)$. In the case $\mathbf{X} = Sym_N$, we consider low-rank matrices $X_0 = UU'$ where U is an N by r partial orthogonal matrix in $St(N, r)$, and again the matrix is uniformly distributed.

For measurement matrices A , we use Gaussian random matrices satisfying $A_{i,j} \sim N(0, 1/n)$.

2.2 Convex Optimization

For a given problem instance (y, A) , we attempt to recover the underlying low-rank object X_0 from the measurements y by convex optimization. Each of our choices \mathbf{X} gives rise to an associated optimization problem:

$$(P_{nuc}^{\mathbf{X}}) \quad \min \|X\|_* \text{ subject to } y = A \cdot \text{vec}(X), \quad X \in \mathbf{X}.$$

Here \mathbf{X} is one of these two classes of matrices Mat_N or Sym_N . The two resulting optimization problems can each be reduced to a so-called semidefinite programming problem; see [12, 13].

2.3 Probability of Exact Recovery

Since both the measurement matrix A , and the underlying low-rank object X_0 are random, (y, A) is a random instance for $(P_{nuc}^{\mathbf{X}})$. The probability of exact recovery is defined by

$$\pi(r|n, N, \mathbf{X}) = \text{Prob}\{X_0 \text{ is the unique solution of } (P_{nuc}^{\mathbf{X}})\}.$$

Clearly $0 \leq \pi \leq 1$; for fixed N , π is monotone decreasing in r and monotone increasing in n . Also $\pi(r|n, N, Mat_N) < \pi(r|n, N, Sym_N)$.

2.4 Estimating the Probability of Exact Recovery

Our procedure follows [14, 15]. For a given matrix type \mathbf{X} and rank r we conduct an experiment whose purpose is to estimate $\pi(r|n, N, \mathbf{X})$ using T Monte Carlo trials. In each trial we generate a random instance (y, A) which we supply to a solver for $(P_{nuc}^{\mathbf{X}})$, obtaining the result X_1 . We compare the result X_1 to X_0 . If the relative error $\|X_0 - X_1\|_F / \|X_0\|_F$ is smaller than a numerical tolerance, we declare the recovery a success; if not, we declare it a failure. (In this paper, we used an error tolerance of 0.001.) We thus obtain T binary measurements Y_i indicating success or failure in reconstruction. The empirical success fraction is then calculated as

$$\hat{\pi}(r|n, N, T, \mathbf{X}) = \frac{\#\{\text{successes}\}}{\#\{\text{trials}\}} = \frac{1}{T} \sum_{i=1}^T Y_i.$$

These are the *raw observations* generated by our experiments.

2.5 Asymptotic Phase Transition Hypothesis

Consider a sequence of tuples (r, n, N) with $r/N \rightarrow \rho$ and $n/N \rightarrow \delta$. We assume that there is an asymptotic phase transition curve $\delta^*(\rho|\mathbf{X})$, i.e. a curve obeying

$$\pi(r|n, N, \mathbf{X}) \rightarrow \begin{cases} 1 & \delta < \delta^*(\rho|\mathbf{X}) \\ 0 & \delta > \delta^*(\rho|\mathbf{X}) \end{cases} \quad (3)$$

For many convex optimization problems the existence of such an asymptotic phase transition is rigorously proven; see the Discussion below.

The hypothesis we investigate concerns the value of $\delta^*(\rho|\mathbf{X})$; specifically, whether

$$\delta^*(\rho|\mathbf{X}) = \mathcal{M}(\rho|\mathbf{X}). \quad (4)$$

Here $\mathcal{M}(\rho|Mat)$ (respectively $\mathcal{M}(\rho|Sym)$) is the minimax MSE for SVT for general matrices (respectively, positive definite ones). Formulas for \mathcal{M} were derived by the Authors in [11]; computational details are provided in the Appendix.

2.6 Empirical Phase Transitions

The *empirical phase transition* point is estimated by fitting a smooth function $\hat{\pi}(n/N)$ (in fact a logistic function) to the empirical data $\hat{\pi}(r|n, N, \mathbf{X})$ using the `glm()` command in the R statistics language. In fact we fit the logistic model that $\text{logit}(\pi) \equiv \log(\frac{\pi}{1-\pi}) = a + b\Delta$, where $\Delta(\delta|\rho) = \delta - M(\rho)$ is the offset between δ and the predicted phase transition. The coefficients a and b are called the intercept and slope, and will be tabulated below. The intercept gives the predicted logit exactly at $\Delta = 0$, i.e. $\delta = \mathcal{M}(\rho)$. The empirical phase transition is located at $\hat{\delta}(r, N, M, \mathbf{X}) = \mathcal{M}(\rho) - a/b$. This is the value of $\delta = \delta(n, N|\mathbf{X})$ solving

$$\hat{\pi}(\delta) = 1/2.$$

Under the hypothesis (4) we have

$$\lim_{N \rightarrow \infty, r/N \rightarrow \rho} \lim_{T \rightarrow \infty} \hat{\delta}(r, N, T, \mathbf{X}) = \mathcal{M}(\rho|\mathbf{X}).$$

Consequently, in data analysis we will compare the fitted values $\hat{\delta}(r, N, T, \mathbf{X})$ with $\mathcal{M}(r/N|\mathbf{X})$.

2.7 Experimental Design

To address our main hypothesis regarding the agreement of phase transition boundaries, we measure $\hat{\pi}$ at points $\delta = n/N$ and $\rho = r/N$ in the phase

plane (δ, ρ) which we expect to be maximally informative about the location of the phase transition. In fact the informative locations in binomial response models correspond to points where the probability of response is in the middle range $(1/10, 9/10)$ [16]. As a rough approximation to such an optimal design, we sample at equispaced $\delta \in [\mathcal{M}(\rho|\mathbf{X}) - 0.05, \mathcal{M}(\rho|\mathbf{X}) + 0.05]$.

2.8 Computing

We primarily used the MATLAB computing environment, and the popular *CVX* convex optimization package [17]. A modeling system for disciplined convex programming by Boyd, Grant and others, supporting two open source interior-point solvers: SeDuMi and SDPT3 [18, 19].

We also studied the robustness of our results across solvers. Zulfikar Ahmed translated our code into Python and used the general purpose solver package CVXOPT by Anderson and Vandeberghe [20].

3 Results

Our experimental data have been deposited at [21] they are contained in a text file with more than 100,000 lines, each line reporting one batch of Monte Carlo experiments at a given r, n, M, N and \mathbf{X} . Each line also documents the number of Monte Carlo trials T , and the observed success fraction $\hat{\pi}$. The file also contains metadata identifying the solver and the researcher responsible for the run.

In all cases, we observed a transition from no observed successes at $\delta = \mathcal{M}(\rho) - 0.05$ to no observed failures at $\delta = \mathcal{M}(\rho) + 0.05$. Figure 3 shows results we obtained at non square matrix ensembles, with varying $\beta = M/N$. The minimax MSE curves $M(\rho|\beta)$ vary widely, but the observed PT's track the curves closely.

Figure 4 shows a small subset of our results in the square case $\mathbf{X} = \text{Mat}_N$, to explain our empirical results; the full tables are given in SI for the square, non square and symmetric positive definite cases. In the square case, the empirical phase transition agrees in all cases with the formula $\mathcal{M}(\rho)$ to two digits accuracy. Table 7 shows that, in the symmetric nonnegative-definite case $\mathbf{X} = \text{Sym}_N$, the empirical phase transition falls within $[\mathcal{M}(\rho) - 0.01, \mathcal{M}(\rho) + 0.01]$.

Previous empirical studies of phase transition behavior in sparse recovery show that, even in cases where the asymptotic phase transition curve is rigorously proven and analytically available, such large- N theory cannot be expected to match empirical finite- N data to within the usual naive standard

ρ	N	T	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	a	b	Z
1/10	40	400	0.351	0.352	-0.128	182.978	-0.581
	50	400	0.351	0.350	0.282	200.131	1.213
	60	400	0.351	0.352	-0.096	221.212	-0.398
	80	400	0.351	0.350	0.415	295.049	1.452
	100	400	0.351	0.349	0.641	383.493	1.900

Figure 4: Results with $\mathbf{X} = \text{Mat}_N$, with $\rho = 1/10$ and Varying N . T total number of reconstructions; a , b : fitted logistic regression parameters; Z : traditional Z-score of logistic regression intercept. See Table 6 for the complete table.

errors [14, 15]. Instead, one observes a finite transition zone of width $\approx c_1/N^{1/2}$ and a small displacement of the empirical phase transition away from the asymptotic Gaussian phase transition, of size $\approx c_2/N$. Hence, the strict literal device of testing the hypothesis that $E\hat{\delta} = \mathcal{M}(\rho)$ is not appropriate in this setting ¹

A precise statement of our hypothesis uses the fitted logistic parameters $\hat{a} = \hat{a}(r, N, M, \mathbf{X})$ and $\hat{b} = \hat{b}(r, N, T, \mathbf{X})$, defined above. The asymptotic relation ²

$$\lim_{N \rightarrow \infty} \lim_{T \rightarrow \infty} \frac{\hat{a}(r, N, T, \mathbf{X})}{\hat{b}(r, N, T, \mathbf{X})} =_P 0, \quad r = \lfloor \rho N \rfloor, \quad (5)$$

implies $\delta(\rho|\mathbf{X}) = \mathcal{M}(\rho|\mathbf{X})$. Now note that in Figure 4 the coefficient b scales directly with N and takes the value several hundred for large N . This means that, in these experiments, the transition from complete failure to complete success happens in a zone of width $< 1/100$. Notice also that a stays bounded, for the most part between 0 and 2. This means that the response probability *evaluated exactly at* $\mathcal{M}(\rho)$ obeys typically:

$$\text{logit}(\hat{\pi}(\mathcal{M}(\rho))) \in [0, 2],$$

Figure 5, Panel A presents the fitted slopes b and problem sizes N , as well as an empirical fit. All the data came from Table 6, but we omitted results for $\rho \in \{3/4, 9/10\}$ because those slopes were large multiples of all other

¹As shown in Figure 4, and in Tables 6, 7, 8, 9, and 10, our results in many cases do generate Z scores for the intercept term in the logistic regression which are consistent with traditional acceptance of this hypothesis. However, traditional acceptance is not needed, in order for the main hypothesis to be valid, and because of finite- N scaling effects indicated above, would not ordinarily be expected to hold.

² $=_P$ denotes convergence in probability.

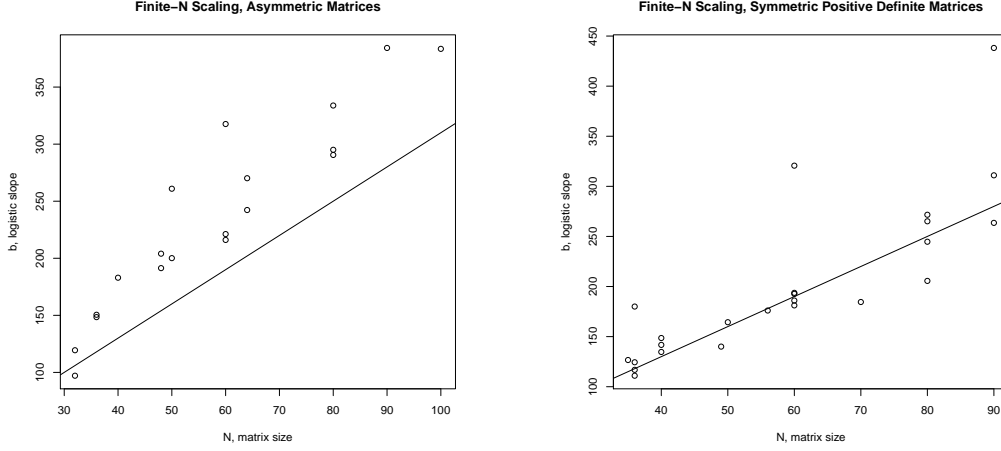


Figure 5: Finite- N scaling effects. Fitted slopes b from Tables 6 and 7 versus matrix sizes N . Left Panel: asymmetric square matrices $\mathbf{X} = \text{Mat}_N$. Right Panel: symmetric nonnegative-definite matrices $\mathbf{X} = \text{Sym}_N$. Superimposed lines have formulas $b = 10 + 3N$. Note: in Left Panel, data for $\rho \in \{3/4, 9/10\}$ were excluded because the slopes b were very large (700 and 2700, respectively).

slopes. Similarly Figure 5, Panel B presents the slopes from Table 7. In each plot, the line $b = 6 + 3N$ is overlaid for comparison. Both plots show a clear tendency of the slopes to increase with N .

The combination of linear growth of b with N and non-growth of a ³ implies Eq. (5), and acceptance of our main hypothesis.

Nongaussian Measurements. This paper studies matrix recovery from Gaussian measurements of the unknown matrix, and specifically does not study recovery from partial entry wise measurements typically called ‘matrix completion’. Entrywise measurements is yield a phase transition at a different location [9]. Our conclusions *do* extend to certain nonGaussian measurements based on \mathcal{A} with independent and identically distributed entries that are equiprobable ± 1 (a.k.a. Rademacher matrices). See Table 10.

³Actually, even sub linear growth of a implies the result.

4 Discussion: Existing Literature and Our Contributions

Phase transitions in the success of convex optimization at solving non-convex problems have been observed previously. Donoho and Tanner considered linear programs useful in the recovery of sparse vectors, rigorously established the existence of asymptotic phase transitions, and derived exact formulas for the asymptotic phase transition [14, 22–25] as well as finite N exponential bounds. Their work considered the recovery of k -sparse vectors with N entries from n Gaussian measurements. The phase diagram in those papers could be stated in terms of variables (κ, δ) , where, as in this paper, $\delta = n/N$ is the under sampling fraction, and $\kappa = k/N$ is the sparsity fraction, which plays a role analogous to the role played here by the rank fraction ρ . The proofs in those papers were obtained by techniques from high-dimensional combinatorial geometry, and the formulas for the asymptotic phase transition were implicit, written in terms of special functions.

Donoho, Maleki, and Montanari [26] later developed a new so-called Approximate Message Passing (AMP) approach to the sparse recovery problem, which gave new formulas for phase boundaries, confirmed rigorously by Bayati and Montanari in [27, 28]. While the previous formulas involved combinatorial geometry, the new (necessarily equivalent) ones involved instead minimax decision theory. An extensive literature on AMP algorithms has since developed see, e.g. [29, 30], implying, among other things, universality in the sparse recovery phase transition [31].

Donoho, Johnstone and Montanari [32] generalized previous AMP-based phase transition results to block-sparse, monotone, and piecewise constant signals. They provided evidence that the observed phase transition $\delta^*(\cdot)$ of the associated convex optimization reconstruction methods obeyed formulas of the form

$$\delta^*(\kappa) = \mathcal{M}(\kappa), \quad 0 < \kappa < 1. \quad (6)$$

Here κ is a variable measuring generalized sparsity and $\mathcal{M}(\kappa)$ the minimax MSE of an appropriate denoiser based on direct noisy measurements. The main result in this brief report fits in this general framework whereby the sparsity κ is identified with the fractional rank ρ , and the minimax MSE symbol \mathcal{M} applies to the singular value thresholding denoiser. Our main result is therefore an extension of DJM-style formulas from the sparse recovery setting problem to the low rank recovery setting.

Much mathematical study of matrix recovery [1, 3, 8, 33] has focused on providing rigorous bounds which show the existence of a region of success, without however establishing a phase transition phenomenon, or determining

its exact boundary. A relatively accurate result by Candès and Recht (CR) for the case of Gaussian measurements [34], implies that $n \geq 3r(M + N)[1 + O(r/N)]$ measurements are sufficient. Our formulas for the square case $M = N$ show that

$$\mathcal{M}(\rho|Mat_N) \sim 6\rho, \quad \rho \rightarrow 0.$$

This agrees with the CR bound in the very low-rank case. However, our relation $\delta^*(\rho) = \mathcal{M}(\rho)$ is apparently noticeably more accurate than the CR formula at finite N . Table 9 presents experiments where the rank is fixed at $r = 1, 2, 3$, or 4, and N varies between 40 and 90. Even though in such cases the corresponding $\rho = r/N$ is rather small, for example $1/90$ in the case $r = 1$ and $N = 90$, the empirical PT in our experiments agrees much more closely with the nonlinear formula $\mathcal{M}(\rho)$ than it does with the linear formula 6ρ . Also, in the non square case $\beta \neq 1$, $\mathcal{M} \sim 2\rho(1 + \beta + \sqrt{\beta})$, which is strictly smaller than 6ρ for $\beta < 1$, so the CR formula is noticeably less accurate than the $\delta^* = \mathcal{M}$ formula in the non square case.

Of the mathematical methods developed to identify phase transitions but which are not based on combinatorial geometry or approximate message passing, the most precise are based on the ‘Escape Through the Mesh’ (ETM) technique of Yoram Gordon [10, 35]. ETM was used to prove upper bounds on the number of Gaussian measurements needed for reconstruction of sparse signals by Stojnic [35] and for low-rank matrices by Oymak and Hassibi [10]. In particular, [10] studies one of the two cases studied here and observes in passing that in the square case, ETM gives bounds that seemingly agree with actual phase transition measurements. Very recently, building on the same approach, a DJM-style inequality $\delta^*(\kappa) \leq \mathcal{M}(\kappa)$ has been announced by Oymak and Hassibi for a wide range of convex optimization problems in [36], including nuclear norm minimization; [36] also presented empirical evidence for $\delta^*(\rho) = \mathcal{M}(\rho)$ in the square case $\mathbf{X} = Mat_N$.

Our Contributions. This paper presents explicit formulas for the minimax MSE of singular value thresholding in various cases, and shows that the new formulas match the appropriate empirical phase transitions in a formal comparison. Compared to earlier work, we make here the following contributions:

- *A Broad Range of Phase Transition Studies* for non square, square, and symmetric nonnegative-definite matrix recovery from Gaussian measurements. We also made certain nonGaussian measurements and observed similar results.
- *A Broad Range of Prediction formulas.* We make available explicit formulas for Minimax MSE in the square symmetric nonnegative-definite,

or asymmetric case, as well as non square cases.

- *Careful Empirical Technique*, including the following:

Reproducibility. Publication of the code and data underlying our conclusions.

Validation. Matlab/CVX results were re-implemented in Python / CVXOPT, with similar results. Code was executed on 3 different computing clusters, with similar results.

Study of Finite- N -scaling. We studied tendencies of a, b as N varies, and observed behavior consistent with our asymptotic phase transition hypothesis. Studies at a single N could only have shown that an empirical phase transition was near to a given theoretical curve, at a given N , but not shown the scaling behavior with N that the main hypothesis properly involves.

5 Conclusions

For the problem of matrix recovery from Gaussian measurements, our experiments, as well as those of others, document the existence of a finite- N phase transition. We compared our measured empirical phase transition curve with a formula from the theory of matrix denoising and observed a compelling match. Although the matrix recovery and matrix denoising problems are superficially different, this match evidences a deeper connection, such that mean squared error properties of a denoiser in a noise removal problem give precisely the exact recovery properties of a matrix recovery rule in a noiseless, but incomplete data problem.

This connection suggests both new limits on what is possible in the matrix recovery problem, but also new ways of trying to reach those limits.

Acknowledgments

Thanks to Iain Johnstone for advice at several crucial points. This work was partially supported by NSF DMS 0906812 (ARRA). MG was supported by a William R. and Sara Hart Kimball Stanford Graduate Fellowship and a Technion EE Sohnis Promising Scientist Award. AM was partially supported by the NSF CAREER award CCF-0743978 and the grant AFOSR/DARPA FA9550-12-1-0411.

References

- [1] Emmanuel J Candès and Benjamin Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2008. URL <http://arxiv.org/abs/0805.4471>.
- [2] David Gross. Recovering Low-Rank Matrices From Few Coefficients in Any Basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, March 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2104999. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5714248>.
- [3] Benjamin Recht, Maryam Fazel, and PA Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010. URL <http://epubs.siam.org/doi/abs/10.1137/070697835>.
- [4] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust Principal Component Analysis ? 2009. URL <http://arxiv.org/pdf/0912.3599>.
- [5] David Gross, Yi-Kai Liu, Steven Flammia, Stephen Becker, and Jens Eisert. Quantum State Tomography via Compressed Sensing. *Physical Review Letters*, 105(15):1–4, October 2010. ISSN 0031-9007. doi: 10.1103/PhysRevLett.105.150401. URL <http://link.aps.org/doi/10.1103/PhysRevLett.105.150401>.
- [6] RH Keshavan, Andrea Montanari, and S Oh. Matrix completion from noisy entries. *The Journal of Machine Learning*, 11:2057–2078, 2010. URL <http://dl.acm.org/citation.cfm?id=1859890.1859920>.
- [7] Benjamin Recht, W Xu, and Babak Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *Proceedings of the 47th IEEE conference on Decision and Control Cancun, Mexico*, 2008. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4739332.
- [8] Benjamin Recht, Weiyu Xu, and Babak Hassibi. Null space conditions and thresholds for rank minimization. *Mathematical Programming*, 127(1):175–202, October 2010. ISSN 0025-5610. doi: 10.1007/s10107-010-0422-2. URL <http://www.springerlink.com/index/10.1007/s10107-010-0422-2>.

- [9] Jared Tanner and KE Wei. Normalized iterative hard thresholding for matrix completion. 2012. URL http://people.maths.ox.ac.uk/tanner/papers/TaWei_NIHT.pdf.
- [10] Samet Oymak and Babak Hassibi. New Null Space Results and Recovery Thresholds for Matrix Rank Minimization. page 28, November 2010. URL <http://arxiv.org/abs/1011.6326>.
- [11] David L. Donoho and Matan Gavish. Minimax Risk of Matrix Denoising by Singular Value Thresholding. Technical report, Stanford University Department of Statistics, 2013.
- [12] Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002. URL <http://www.inatel.br/docentes/dayan/TP524/Artigos/MatrixRankMinimizationwithApplications.pdf>.
- [13] Maryam Fazel, H Hindi, and Stephen P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference, 2001*, pages 4734–4739, 2001. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=945730.
- [14] David L. Donoho and Jared Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical transactions of the Royal Society, Series A*, 367(1906):4273–93, November 2009. ISSN 1364-503X. doi: 10.1098/rsta.2009.0152. URL <http://www.ncbi.nlm.nih.gov/pubmed/19805445>.
- [15] Hatef Monajemi, Sina Jafarpour, Matan Gavish, and David L. Donoho. Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):1181–6, January 2013. ISSN 1091-6490. doi: 10.1073/pnas.1219540110. URL <http://www.pnas.org/cgi/content/long/110/4/1181>.
- [16] Leslie A. Kalish. Efficient design for estimation of median lethal dose and quantal dose-response curves. *Biometrics*, 46(3):737–48, September 1990. ISSN 0006-341X. URL <http://www.ncbi.nlm.nih.gov/pubmed/2242412>.
- [17] M. Grant and Stephen P. Boyd. CVX: Matlab Software for Disciplined Convex Programming, 2010. URL cvxr.com/cvx.

- [18] JF Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization methods and software*, 1999. URL <http://www.tandfonline.com/doi/abs/10.1080/10556789908805766>.
- [19] K.C. Toh, M.J. Todd, and Tutuncu R.H. SDPT3 - a Matlab software package for semidefinite programming, *Optimization Methods and Software*, 1999.
- [20] D. Andersen, J. Dahl, and L. Vandenberghe. CVXOPT: Python Software for Convex Optimization, 2012. URL <http://abel.ee.ucla.edu/cvxopt>.
- [21] David L. Donoho, Matan Gavish, and Andrea Montanari. Data for the article The Phase Transition of Matrix Recovery from Gaussian Measurements Matches the Minimax MSE of Matrix Denoising. <http://purl.stanford.edu/tz124hw0000>, 2013. URL purl.stanford.edu/tz124hw0000. accessed 15 February 2013.
- [22] David L. Donoho. High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension. *Discrete & Computational Geometry*, 35(4):617–652, December 2005. ISSN 0179-5376. doi: 10.1007/s00454-005-1220-0. URL <http://www.springerlink.com/index/10.1007/s00454-005-1220-0>.
- [23] David L. Donoho and Jared Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.
- [24] David L. Donoho and Jared Tanner. Precise Undersampling Theorems. *Proceedings of the IEEE*, 98(6):913–924, June 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2010.2045630. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5458001>.
- [25] David L. Donoho and Jared Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9452–7, July 2005. ISSN 0027-8424. doi: 10.1073/pnas.0502258102. URL <http://www.pnas.org/cgi/content/long/102/27/9452>.
- [26] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45):18914–9,

- November 2009. ISSN 1091-6490. doi: 10.1073/pnas.0909892106. URL <http://www.pnas.org/cgi/content/long/106/45/18914>.
- [27] Mohsen Bayati and Andrea Montanari. The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, February 2011. ISSN 0018-9448. doi: 10.1109/TIT.2010.2094817. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5695122>.
 - [28] Mohsen Bayati and Andrea Montanari. The LASSO Risk for Gaussian Matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, April 2012. ISSN 0018-9448. doi: 10.1109/TIT.2011.2174612. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6069859>.
 - [29] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. Ieee, July 2011. ISBN 978-1-4577-0596-0. doi: 10.1109/ISIT.2011.6033942. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6033942>.
 - [30] Jeremy Vila and Philip Schniter. Expectation-maximization Bernoulli-Gaussian approximate message passing. In *Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2011*, number July, pages 799–803, 2011. ISBN 9781467303231. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6190117.
 - [31] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in Polytope Phase Transitions and Message Passing Algorithms. July 2012. URL <http://arxiv.org/abs/1207.7321>.
 - [32] David L. Donoho, Iain M Johnstone, and Andrea Montanari. Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising. 2011. URL <http://http://arxiv.org/abs/1111.1041v2>.
 - [33] Samet Oymak, Karthik Mohan, Maryam Fazel, and Babak Hassibi. A simplified approach to recovery conditions for low rank matrices. *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2318–2322, July 2011. doi: 10.1109/ISIT.

- 2011.6033976. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6033976>.
- [34] Emmanuel J Candès and Benjamin Recht. Simple bounds for recovering low-complexity models. *Mathematical Programming*, (June 2011):1–11, 2011. URL <http://www.springerlink.com/index/655751h7568x3621.pdf>.
- [35] Mihailo Stojnic. Various thresholds for l1-optimization in compressed sensing. July 2009. URL <http://arxiv.org/abs/0907.3666>.
- [36] Samet Oymak and Babak Hassibi. On a Relation between the Minimax Denoising and the Phase Transitions of Convex Functions, 2012. URL <http://www.its.caltech.edu/~soymak/Relation.pdf>.
- [37] VA Marcenko and LA Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics USSR Sbornik*, 1(4):457–483, 1967. URL <http://iopscience.iop.org/0025-5734/1/4/A01>.

A Asymptotic Minimax MSE Formula

The following provides explicit formulas for the matrix denoising minimax curves $\mathcal{M}(\rho, \beta | Mat)$ and $\mathcal{M}(\rho | Sym)$ used above. Please see [11] for the derivations. Computer programs that efficiently calculate these quantities are provided in [21]. Let

$$P_\gamma(x; k) = \frac{1}{2\pi\gamma} \int_x^{\gamma_+} t^{k-1} \sqrt{(\gamma_+ - t)(t - \gamma_-)} dt, \quad (7)$$

where $\gamma_\pm = (1 \pm \sqrt{\gamma})^2$, denote the complementary incomplete moments of the Marčenko-Pastur distribution [37]. Define

$$\begin{aligned} \mathbf{M}_\alpha(\Lambda; \rho, \tilde{\rho}) = & \rho + \tilde{\rho} - \rho\tilde{\rho} + (1 - \tilde{\rho}) \left[\rho\Lambda^2 + \right. \\ & \left. + \alpha(1 - \rho) \left(P_\gamma(\Lambda^2; 1) - 2\Lambda P_\gamma(\Lambda^2; \tfrac{1}{2}) + \Lambda^2 P_\gamma(\Lambda^2; 0) \right) \right], \end{aligned} \quad (8)$$

with $\gamma = \gamma(\rho, \tilde{\rho}) = (\tilde{\rho} - \rho\tilde{\rho})/(\rho - \rho\tilde{\rho})$.

Case $\mathbf{X} = \text{Mat}_{M,N}$. The minimax curve is given by $\mathcal{M}(\rho, \beta | \text{Mat}) = \inf_{\Lambda} \mathbf{M}_1(\Lambda; \rho, \beta \rho)$. The following minimaxity interpretation is proved in [11]

$$\lim_{N \rightarrow \infty} \inf_{\lambda} \sup_{\substack{X_0 \in \mathbb{R}^{\lfloor \beta N \rfloor \times N} \\ \text{rank}(X_0) \leq \rho \beta N}} \text{MSE}(\hat{X}_{\lambda}, X_0) = \mathcal{M}(\rho, \beta | \text{Mat}). \quad (9)$$

Case $\mathbf{X} = \text{Sym}_N$. The minimax curve is given by $\mathcal{M}(\beta | \text{Sym}) = \inf_{\Lambda} \mathbf{M}_{1/2}(\Lambda; \rho, \rho)$. The following minimaxity interpretation is proved in [11]

$$\lim_{N \rightarrow \infty} \inf_{\lambda} \sup_{\substack{X_0 \in S_N \\ \text{rank}(X_0) \leq \rho N}} \text{MSE}(\hat{X}_{\lambda}, X_0) = \mathcal{M}(\rho | \text{Sym}). \quad (10)$$

Computing $\mathcal{M}(\rho, \beta | \mathbf{X})$. The map $\Lambda \mapsto \mathbf{M}_{\alpha}(\Lambda; \rho, \tilde{\rho})$ is convex. Solving $d\mathbf{M}_{\alpha}/d\Lambda = 0$ we get that $\text{argmin}_{\Lambda} \mathbf{M}_{\alpha}(\Lambda; \rho, \tilde{\rho})$ is the unique root of the equation

$$\Lambda^{-1} P_{\gamma}(\Lambda^2; \frac{1}{2}) - P_{\gamma}(\Lambda^2; 0) = \frac{\rho}{\alpha(1 - \rho)}. \quad (11)$$

The right hand side of (11) is decreasing in Λ and the solution is determined numerically by binary search.

For square matrices ($\rho = \tilde{\rho}$) (8) can be expressed using elementary trigonometric functions. In [11] it is shown that

$$\begin{aligned} \mathbf{M}_{\alpha}(\Lambda; \rho, \rho) &= \rho(2 - \rho) + (1 - \rho) [\rho \Lambda^2 + \\ &+ \alpha(1 - \rho) (Q_2(\Lambda) - 2\lambda Q_1(\Lambda) + \Lambda^2 Q_0(\Lambda))] \end{aligned} \quad (12)$$

where

$$\begin{aligned} Q_0(x) &= \frac{1}{\pi} \int_x^2 \sqrt{4 - x^2} = 1 - \frac{x}{2\pi} \sqrt{4 - x^2} - \frac{2}{\pi} \text{atan}\left(\frac{x}{\sqrt{4 - x^2}}\right) \\ Q_1(x) &= \frac{1}{\pi} \int_x^2 x \sqrt{4 - x^2} = \frac{1}{3\pi} (4 - x^2)^{3/2} \\ Q_2(x) &= \frac{1}{\pi} \int_x^2 x^2 \sqrt{4 - x^2} = 1 - \frac{1}{4\pi} x \sqrt{4 - x^2} (x^2 - 2) - \frac{2}{\pi} \text{asin}\left(\frac{x}{2}\right) \end{aligned}$$

are the complementary incomplete moments of the Quarter Circle law. Moreover

$$\text{argmin}_{\Lambda} \mathbf{M}_{\alpha}(\Lambda; \rho, \rho) = 2 \cdot \sin(\theta_{\alpha}(\rho)), \quad (13)$$

where $\theta_{\alpha}(\rho) \in [0, \pi/2]$ is the unique solution to the transcendental equation

$$\theta + \cot(\theta) \cdot \left(1 - \frac{1}{3} \cos^2(\theta)\right) = \frac{\pi(1 + \alpha^{-1}\rho - \rho)}{2(1 - \rho)}, \quad (14)$$

which is a simplified version of (11).

Parametric representation of the minimax curves. For square matrices ($\rho = \tilde{\rho}$) the minimax curves $\mathcal{M}(\rho, 1|Mat)$ and $\mathcal{M}(\rho|Sym)$ admit a parametric representation in the (ρ, \mathcal{M}) plane using elementary trigonometric functions, see [11]. As θ ranges over $[0, \pi/2]$,

$$\begin{aligned}\rho(\theta) &= 1 - \frac{\pi/2}{\theta + (\cot(\theta) \cdot (1 - \frac{1}{3}\cos^2(\theta)))} \\ \mathcal{M}(\theta) &= 2\rho(\theta) - \rho^2(\theta) + 4\rho(\theta)(1 - \rho(\theta))\sin^2(\theta) \\ &+ \frac{4}{\pi}(1 - \rho)^2 \left[(\pi - 2\theta)\left(\frac{5}{4} - \cos(\theta)^2\right) + \frac{\sin(2\theta)}{12}(\cos(2\theta) - 14) \right]\end{aligned}$$

is a parametric representation of $\mathcal{M}(\rho, 1|Mat)$, and similarly

$$\begin{aligned}\rho(\theta) &= 1 - \frac{\theta + (\cot(\theta) \cdot (1 - \frac{1}{3}\cos^2(\theta))) - \pi/2}{\theta + (\cot(\theta) \cdot (1 - \frac{1}{3}\cos^2(\theta))) + \pi/2} \\ \mathcal{M}(\theta) &= 2\rho(\theta) - \rho^2(\theta) + 4\rho(\theta)(1 - \rho(\theta))\sin^2(\theta) \\ &+ \frac{2}{\pi}(1 - \rho)^2 \left[(\pi - 2\theta)\left(\frac{5}{4} - \cos(\theta)^2\right) + \frac{\sin(2\theta)}{12}(\cos(2\theta) - 14) \right]\end{aligned}$$

is a parametric representation of $\mathcal{M}(\rho|Sym)$.

B Summary of Empirical Results

ρ	N	T	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	a	b	Z
1/10	40	400	0.351	0.352	-0.128	182.978	-0.581
	50	400	0.351	0.350	0.282	200.131	1.213
	60	400	0.351	0.352	-0.096	221.212	-0.398
	80	400	0.351	0.350	0.415	295.049	1.452
	100	400	0.351	0.349	0.641	383.493	1.900
1/8	32	400	0.414	0.412	0.262	119.424	1.457
	48	400	0.414	0.413	0.262	204.025	1.120
	64	400	0.414	0.412	0.560	270.147	2.006
	80	400	0.414	0.413	0.296	290.536	1.057
1/6	36	400	0.507	0.505	0.356	150.509	1.749
	60	400	0.507	0.506	0.137	216.067	0.571
1/4	32	400	0.655	0.651	0.398	97.107	2.375
	48	400	0.655	0.653	0.356	191.373	1.554
	64	400	0.655	0.653	0.507	242.289	1.935
	80	400	0.655	0.651	1.312	333.842	3.584
1/3	36	400	0.765	0.765	-0.006	148.439	-0.029
	60	400	0.765	0.762	1.145	317.633	3.348
	90	400	0.765	0.762	1.487	384.289	3.610
1/2	50	400	0.905	0.903	0.658	260.939	2.361
3/4	40	400	0.989	0.986	1.535	709.400	3.837
9/10	50	400	0.999	0.998	2.478	2980.216	3.435

Figure 6: Results with $\mathbf{X} = Mat_N$.

ρ	N	T	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	a	b	Z
1/10	40	800	0.315	0.310	0.787	148.605	4.268
	60	800	0.315	0.312	0.519	186.020	2.640
	80	800	0.315	0.311	0.995	265.114	3.864
1/8	40	800	0.371	0.369	0.270	141.743	1.585
	56	800	0.371	0.369	0.230	176.063	1.212
	80	800	0.371	0.369	0.368	244.721	1.628
1/7	35	800	0.407	0.403	0.623	126.659	3.730
	49	800	0.407	0.404	0.496	140.034	2.878
	70	800	0.407	0.404	0.581	184.556	2.898
1/6	36	800	0.453	0.447	0.702	116.847	4.607
	60	800	0.453	0.447	1.097	193.706	5.133
	90	800	0.453	0.449	1.170	311.016	4.246
1/5	40	800	0.511	0.505	0.786	134.698	5.448
	50	800	0.511	0.507	0.707	164.398	4.528
	80	800	0.511	0.507	0.859	205.621	4.778
1/4	36	800	0.588	0.579	0.967	110.989	7.120
	60	800	0.588	0.582	1.155	181.174	6.379
	80	800	0.588	0.583	1.450	271.651	6.066
1/3	36	800	0.694	0.685	1.085	124.413	3.926
	60	800	0.694	0.688	1.034	192.690	5.710
	90	800	0.694	0.689	1.212	263.543	5.481
1/2	36	800	0.844	0.837	1.306	180.027	6.965
	60	800	0.844	0.838	1.872	320.651	6.425
	90	800	0.844	0.840	1.884	438.138	5.513

Figure 7: Results with $\mathbf{X} = Sym_N$.

β	ρ	$\mathcal{M}(\rho)$	$\delta(\rho)$	a	b	Z	$\sqrt{M \cdot N}$
1/4	0.100	0.241	0.241	0.020	215.506	0.085	60.000
	0.125	0.290	0.290	-0.037	322.468	-0.127	64.000
	0.143	0.323	0.321	0.468	238.477	1.813	70.000
	0.167	0.365	0.362	0.884	247.690	3.109	60.000
	0.200	0.421	0.422	-0.153	246.541	-0.597	60.000
	0.250	0.498	0.496	0.339	214.134	1.405	64.000
	0.333	0.610	0.607	0.856	284.629	2.834	60.000
	0.500	0.788	0.786	0.635	297.803	2.124	64.000
1/3	0.100	0.255	0.255	-0.105	269.362	-0.395	51.962
	0.143	0.340	0.340	0.031	210.274	0.133	48.497
	0.167	0.383	0.381	0.373	169.227	1.733	51.962
	0.200	0.440	0.437	0.589	197.335	2.457	51.962
	0.250	0.517	0.517	0.113	218.479	0.470	55.426
	0.333	0.629	0.626	0.661	219.930	2.584	51.962
	0.500	0.801	0.799	0.572	274.122	2.036	55.426
1/2	0.200	0.475	0.474	0.127	159.697	0.616	42.426
	0.250	0.554	0.553	0.258	201.035	1.114	45.255
	0.286	0.604	0.603	0.307	167.955	1.434	49.497
	0.333	0.665	0.662	0.486	155.704	2.327	42.426
	0.400	0.738	0.737	0.292	184.091	1.308	42.426
	0.500	0.827	0.825	0.460	228.287	1.826	45.255
	0.667	0.930	0.927	0.729	248.019	3.073	42.426
3/5	0.100	0.296	0.295	0.146	191.950	0.646	38.730
	0.125	0.352	0.351	0.242	219.767	1.000	61.968
	0.143	0.389	0.388	0.105	179.344	0.484	54.222
	0.167	0.436	0.433	0.425	198.788	1.814	46.476
	0.200	0.495	0.494	0.237	146.409	1.196	38.730
	0.250	0.575	0.573	0.361	164.339	1.704	46.476
	0.333	0.686	0.684	0.509	258.045	1.886	46.476
	0.500	0.843	0.842	0.153	156.137	0.751	46.476
2/3	0.100	0.305	0.306	-0.168	221.047	-0.697	36.742
	0.125	0.363	0.362	0.064	151.402	0.321	39.192
	0.143	0.401	0.399	0.263	142.052	1.346	34.293
	0.167	0.448	0.446	0.271	157.783	1.323	36.742
	0.200	0.509	0.510	-0.209	155.201	-1.034	36.742
	0.250	0.589	0.587	0.315	152.493	1.555	39.192
	0.333	0.700	0.697	0.408	190.613	1.768	36.742
	0.500	0.854	0.853	0.257	232.865	1.034	39.192
3/4	0.100	0.317	0.317	-0.043	163.597	-0.207	34.641
	0.125	0.376	0.375	0.237	192.933	1.047	55.426
	0.143	0.415	0.412	0.562	178.363	2.476	48.497
	0.167	0.463	0.461	0.304	174.599	1.405	41.569
	0.200	0.525	0.524	0.159	120.459	0.886	34.641
	0.250	0.606	0.604	0.276	154.259	1.353	41.569
	0.333	0.716	0.714	0.378	180.963	1.682	41.569
	0.500	0.867	0.866	0.341	283.911	1.227	41.569
4/5	0.100	0.324	0.323	0.220	174.269	1.019	44.721
	0.125	0.384	0.381	0.423	160.001	1.999	44.721
	0.143	0.423	0.422	0.273	248.331	1.055	62.610
	0.167	0.472	0.472	-0.025	190.094	-0.112	40.249
	0.200	0.534	0.533	0.204	195.806	0.889	44.721
	0.250	0.616	0.616	-0.076	197.138	-0.333	44.721
	0.333	0.726	0.725	0.213	203.942	0.910	40.249
	0.500	0.875	0.873	0.412	215.351	1.689	44.721

Figure 8: Results with non square matrices $\mathbf{X} = \text{Mat}_{M,N}$. Each row based on $T = 400$ Monte Carlo trials.

N	r	ρ	6ρ	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	a	b	Z
90	1	0.011	0.067	0.059	0.054	3.595	828.597	4.627
80	1	0.013	0.075	0.065	0.059	2.766	534.431	5.117
70	1	0.014	0.086	0.072	0.069	1.444	432.528	3.962
60	1	0.017	0.100	0.081	0.078	1.374	434.911	3.542
50	1	0.020	0.120	0.094	0.091	1.159	348.687	3.226
90	2	0.022	0.133	0.103	0.101	1.225	611.400	2.499
40	1	0.025	0.150	0.114	0.114	-0.169	328.547	-0.513
40	1	0.025	0.150	0.114	0.114	-0.169	328.547	-0.513
70	2	0.029	0.171	0.127	0.125	6.383	2746.893	0.007
30	1	0.033	0.200	0.145	0.145	0.013	137.638	0.051
30	1	0.033	0.200	0.145	0.145	0.013	137.638	0.051
30	1	0.033	0.200	0.145	0.145	0.013	137.638	0.051
80	3	0.037	0.225	0.160	0.158	0.493	345.100	1.111
50	2	0.040	0.240	0.169	0.166	1.071	364.190	1.981
70	3	0.043	0.257	0.179	0.177	3.442	1883.417	0.006
90	4	0.044	0.267	0.184	0.183	2.044	1773.906	0.006
20	1	0.050	0.300	0.203	0.203	-0.055	94.130	-0.219
20	1	0.050	0.300	0.203	0.203	-0.055	94.130	-0.219
20	1	0.050	0.300	0.203	0.203	-0.055	94.130	-0.219
20	1	0.050	0.300	0.203	0.203	-0.055	94.130	-0.219
70	4	0.057	0.343	0.226	0.227	-2.755	2732.891	-0.001
50	3	0.060	0.360	0.235	0.230	6.462	1302.371	0.008
30	2	0.067	0.400	0.256	0.255	0.106	152.849	0.290
30	2	0.067	0.400	0.256	0.255	0.106	152.849	0.290
40	3	0.075	0.450	0.281	0.280	0.114	191.765	0.249
50	4	0.080	0.480	0.296	0.294	0.361	225.965	0.622
20	2	0.100	0.600	0.351	0.345	0.387	60.051	1.345
20	2	0.100	0.600	0.351	0.345	0.387	60.051	1.345
20	2	0.100	0.600	0.351	0.345	0.387	60.051	1.345
30	4	0.133	0.800	0.434	0.434	-0.005	115.737	-0.011
20	3	0.150	0.900	0.472	0.480	-0.690	86.275	-1.510
20	4	0.200	1.200	0.572	0.587	-0.994	64.706	-2.057

Figure 9: Results with low rank square matrices $\mathbf{X} = \text{Mat}_{N,N}$. Each row based on $T = 400$ Monte Carlo trials.

ρ	N	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	a	b	Z
1/10	40.000	0.351	0.350	0.187	170.554	0.878
	60.000	0.351	0.350	0.239	286.765	0.863
	80.000	0.351	0.351	0.166	290.560	0.597
1/8	40.000	0.414	0.413	0.145	146.920	0.736
	56.000	0.414	0.412	0.552	215.088	2.220
	80.000	0.414	0.413	0.667	375.205	1.992
1/7	35.000	0.456	0.457	-0.158	141.742	-0.813
	49.000	0.456	0.455	0.158	216.040	0.659
	70.000	0.456	0.454	0.911	394.564	2.527
1/6	36.000	0.507	0.505	0.370	159.357	1.768
	60.000	0.507	0.505	0.570	245.893	2.137
	90.000	0.507	0.505	0.689	335.299	2.168
1/5	40.000	0.572	0.571	0.089	172.047	0.417
	60.000	0.572	0.568	0.912	248.273	3.188
	80.000	0.572	0.570	0.598	285.542	2.073
1/3	36.000	0.765	0.759	0.752	124.479	3.825
	60.000	0.765	0.763	0.608	234.115	2.324
	81.000	0.765	0.762	1.053	288.525	3.306
1/2	36.000	0.905	0.903	0.487	229.398	1.914
	60.000	0.905	0.902	1.403	379.105	3.518
	90.000	0.905	0.903	11.365	3971.012	0.008

Figure 10: Results with Rademacher measurements of square matrices $\mathbf{X} = \text{Mat}_{N,N}$. Each row based on $T = 400$ Monte Carlo trials.

C Data Deposition

The data have been deposited in a text file at [21]. A typical fragment of the file is given here:

```

Line Project      Experiment      M N S Instance rank rho      delta      Err0      Err1 Err2
3381 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.73395061728395 0.016630719182629 0 0.4444444444444444
3382 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.739213775178687 0.0159010475527301 0 0.4652777777777778
3383 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.744476933073424 0.0161172486497232 0 0.4166666666666667
3384 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.749740090968161 0.00149080158529591 1 1
3385 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.755003248862898 0.0340839945130298 0 0.1736111111111111
3386 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.760266406757635 0.0235186056093925 0 0.3611111111111111
3387 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.765529564652372 0.0136400454215757 0 0.5069444444444444
3388 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.770792722547109 1.11644848368808e-09 1 1
3389 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.776055880441845 0.00194118165102407 0 1
3390 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.781319038336582 0.0059943065062774 0 0.9027777777777778
3391 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.786582196231319 1.83878516274726e-09 1 1
3392 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.791845354126056 7.85225405546251e-09 1 1
3393 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.797108512020793 4.47138355029567e-10 1 1
3394 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.80237166991553 1.03566257175607e-08 1 1
3395 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.807634827810266 2.62954112892325e-09 1 1
3396 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.812897985705003 0.00120410713874671 1 1
3397 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.81816114359974 1.03259685844506e-08 1 1
3398 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.823424301494477 2.52202641519081e-10 1 1
3399 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.828687459389214 1.94437637948599e-09 1 1
3400 Nuc_CVX_20121129dii Nuc_CVX_N12S01m 12 12 1 m 4 0.3333333333333333 0.83395061728395 1.69829115472996e-10 1 1

```

The fields have the following meaning

- Line – Line number in file; in the above example, lines 3381-3400.
- Project – File identifier – allows identification of code and logs that generated these data; in the above fragment, 'Nuc_CVX_20121129dii'.
- Experiment – File identified – allows identification of code and logs that generated these data; in the above fragment 'Nuc_CVX_N12S01m'.
- M,N – matrix size of X_0 , i.e M by N matrix; in the above fragment $M = N = 12$.
- S – number of matrices in a stack (see below)
- Instance – alphabetic code a-t, identifying one of 20 identical runs which generated this result; in the above fragment, 'm'.
- rank – integer rank of matrix; in the above fragment 'm'.
- rho – fraction in $[0, 1]$, $\rho = \text{rank}/N$; in the above fragment, $1/3$.
- delta – $\delta = n/(MN)$ in case of asymmetric matrix, or $\delta = 2*n/(N(N+1))$ in case of symmetric matrix.
- Err0 – $\|\hat{X} - X_0\|_F/(NM)^{1/2}$.
- Err1 – 1 iff $\|\hat{X} - X_0\|_F/\|X_0\|_F < tol$, and 0 otherwise
- Err2 – fraction of entries with discrepancy $|\hat{X}(i, j) - X_0(i, j)| < tol$.

Additional concepts:

- *Numerical Tolerance.* In our experiments, we used a numerical error tolerance parameter $tol = 0.001$.
- Our experiments also extensively covered cases where X_0 is a 'stack of matrices', i.e. a 3-way array $M \times N \times S$, where S is the number of items in the stack. The only cases of interest for this paper are $S = 1$.

D Code Deposition

There are two types of code deposition.

- *Reproduction from Data Deposition.* The code that actually makes the figures and tables we presented in this paper, starting from the data deposition. This is deposited at [21]. The code we actually ran to create our figures and tables is a set of R scripts, and was run on a Mac OS X. We believe the same code runs with minimal changes on a LINUX environment.
- *RunMyCode Deposition.* For readers who wish simply to compute *the value of the Minimax Mean-Squared Error* over each of the matrix classes we considered, we offer a Minimax MSE calculator at RunMyCode.org.
- *Full Code and Results Deposition.* At [21], we also offer a literal dump of the code we ran and all the logs and result files we obtained.

We believe the first two items are self-documenting. The third item can be explained as follows. Our database of the experiment and all results is contained in a unix directory tree rooted at `exp`.

Inside directory `exp` one finds further directories, as indicated below:

```
Nuc_CVX_20120605a
Nuc_CVX_20120607a
...
Nuc_CVX_20121107a
Nuc_CVX_20121113a
...
Nuc_CVX_20121120a
Nuc_CVX_20121121a
...
SDP_CVX_20121230h
```

```

SDP_CVX_20121230i
..
Nuc_CVX_20130110a
Nuc_CVX_20130110c
...
Nuc_CVX_20130117Cc
Nuc_CVX_20130117Cd
...
Nuc_CVX_20130126Df
Nuc_CVX_20130126Dg
Nuc_CVX_20130126Dh

```

These directory names are precisely the **Project** names used in the data deposition. Say we look inside one of these directories, for example 'Nuc_CVX_20121121b'. We will find a directory called **bin** containing software, and a list of further directories. We excerpt from a 2-column listing of those directories:

```

Nuc_CVX_N05S1a Nuc_CVX_N30S1c
Nuc_CVX_N05S1b Nuc_CVX_N30S1d
Nuc_CVX_N05S1c Nuc_CVX_N30S1e
...
Nuc_CVX_N10S1j Nuc_CVX_N35S1l
Nuc_CVX_N10S1k Nuc_CVX_N35S1m
Nuc_CVX_N10S1l Nuc_CVX_N35S1n
...
Nuc_CVX_N15S1h Nuc_CVX_N40S1j
Nuc_CVX_N15S1i Nuc_CVX_N40S1k
Nuc_CVX_N15S1j Nuc_CVX_N40S1l
...
Nuc_CVX_N20S1o Nuc_CVX_N45S1q
Nuc_CVX_N20S1p Nuc_CVX_N45S1r
Nuc_CVX_N20S1q Nuc_CVX_N45S1s
...
Nuc_CVX_N25S1n Nuc_CVX_N50S1p
Nuc_CVX_N25S1o Nuc_CVX_N50S1q
Nuc_CVX_N25S1p Nuc_CVX_N50S1r
...

```

These directory names are precisely the '**Experiment**' values seen in the data deposition. Inside the **bin** directory we find the matlab code used in common by all the above experiments.

```
Nuc_CVX_20121121b$ ls bin
aveRank.m solveNuc_CVX_Stack_Arb.m
predNucPT.m stackRankRMatrix.m
rankRMatrix.m
```

Inside one of the experiment directories we find experiment-specific files (in both Matlab and Bash script) as well as output files. The subdirectory `logs` contains logs that were created while the jobs were running.

```
Nuc_CVX_20121121b $ ls -R -C1 Nuc_CVX_N25S1o
bashMain.sh
PTExperiment.m
matlabMain.m
results.mat
randState.mat
```

```
Nuc_CVX_N25S1o/logs:
Nuc_CVX_N25S1o.stderr
Nuc_CVX_N25S1o.stdout
runMatlab.20121121b_Nuc_CVX_N25S1o.log
```

Here the `.mat` files were produced by matlab during the running of the experiment.

- `randState.mat` preserves the state of the random number generator at the beginning of that experiment.
- `results.mat` gives the results of the individual problem instances; typically in the form $(r, n, N, M, Err0, Err1, Err2)$.